On Anderson's Probability Inequality

by

Somesh Das Gupta*
University of Minnesota

December 1982

Technical Report No. 413

*Supported by National Science Foundation grant NSF/MCS8202209.

1. Introduction

Anderson's probability inequality [1] has led to a significant development of research on probability inequalities, especially applicable to multivariate inference. The 1955 paper of Anderson has three major facets. Firstly, it introduced a definition of multivariate unimodal function. Secondly, under unimodal probability density, it studied the probability content of a centrally symmetric convex set translated along a ray through the origin. Thirdly, it demonstrated that the convolution of two centrally symmetric unimodal densities in R^n $(n > 1)$ may not be unimodal.

It seems to be appropriate to discuss some modifications, generalizations and consequences of Anderson's inequalities on the occasion of his sixty-fifth birthday in order to indicate the impact of Anderson's contributions. Let us now state Anderson's inequality.

Theorem (Anderson). Let E be a convex set in n-space, symmetric about the origin. Let $f(x) \geqslant 0$ be a function such that (i) $f(x) = f(-x)$ (ii) $\{x | f(x) \geqslant u\} = K_u$ is convex for every $u (0 < u < \infty)$, and (iii) $\int_E f(x)dx < \infty$ (in the Lebesgue sense). Then $\int_E f(x + ky)dx \geqslant \int_E f(x + y)dx$ for $0 \leqslant k \leqslant 1$.

2. Generalizations with Symmetric Functions

First let us indicate the basic steps in the proof of Anderson's inequality. Note that

(2.1) $H(y) \equiv \int_{E+y} f(x)\,dx = \int_0^\infty h(y,u)\,du,$

where

(2.2) $h(y,u) = \int_{R^n} \chi(x;K_u)\ \chi(x;E+y)\,dx,$

and χ stands for indicator function.

An application of Brunn-Minkowski inequality yields

(2.3) $h(\lambda_1 y_1 + \lambda_2 y_2, u) \geqslant \min[h(y_1,u),\ h(y_2,u)],$

where $0 \leqslant \lambda_1, \lambda_2 \leqslant 1$. Specializing $\lambda_1 = (1+\lambda)/2$, $y_1 = y$, $y_2 = -y$, and noting that

(2.4) $h(y,u) = h(-y,u),$

we get

(2.5) $h(\lambda y,u) \geqslant h(y,u).$

The above result implies

(2.6) $H(\lambda y) \geqslant H(y),$

A function H will be called ray-unimodal if it satisfies (2.6).

We may write

(2.7) $H(y) = \int f(x)\,\chi(x-y;E)\,dx$

So H is the convolution of f and $\chi(\cdot,D)$. The first question on generalization considered in the literature was whether the ray-unimodality property is enjoyed by the convolution of more general types of symmetric functions. It follows easily that the convolution of two functions, each of which is a positive mixture of symmetric unimodal functions, is ray-unimodal. Following this line of throught, Sherman [15] has shown that the closed (in the sense of max of L_1-norm and sup-norm) convex cone C_3 generated by indicator functions of symmetric compact convex sets in R^n is closed under convolution. Moreover, any function H in C_3 satifies

$$H(y) = H(-y), \; H(\lambda y) \geq H(y),$$

for $0 \leq \lambda \leq 1$. Since $\int_E f(x+y)dx \in C_3$, Anderson's inequality follows from Sherman's result.

Dharmadhikari and Jogdeo [6] introduced two notions of multivariate unimodality. They called a distribution P on R^n <u>central convex UM</u> if it is the closed (in the sense of weak convergence) convex hull of the set of all uniform distributions on symmetric compact convex bodies in R^n. Moreover, a distribution P on R^n is called <u>monotone UM</u> if for every symmetric convex set C in R^n and every nonzero vector x in R^n, $P(C+kx)$ is nonincreasing in $k \in [0,\infty)$. It follows easily that a central convex UM distribution and a monotone UM distribution is symmetric.

Anderson's result essentially states that every distribution in R^n with symmetric unimodal density is monotone UM. Dharmadhikari and Jogdeo [6] have shown that monotone unimodality is closed under weak convergence. Thus Sherman's result [15] implies that every central convex UM distribution is monotone UM.

It follows trivially that $\int f(x+ky)dP(x)$ is nonincreasing in $k \in [0,\infty)$, where f is a symmetric UM function and P is a monotone UM distribution; this generalization is due to Dharmadhikari and Jogdeo [6].

The basic question relating Anderson's inequalities is regarding the notion of multivariate unimodality. It appears that Anderson's definition is too restrictive. For example the function

$$f(x,y) = \frac{1}{\pi^2} \; \frac{1}{1+x^2} \; \frac{1}{1+y^2}.$$

is not unimodal according to Anderson's definition. Another drawback of Anderson's notion of unimodality is the fact that it is not closed under convolution. This was in fact demonstrated by an example of Anderson [1]. On the other hand, Dharmadhikari and Jogdeo [6] have shown that the convolution between a central convex UM distribution and a monotone UM distribution is monotone UM.

Kanter [10] introduced a more general notion of symmetric unimodal distributions which enjoy many desirable properties. Note that a symmetric unimodal function f on R^n may be expressed as

$$f(x) = \int_0^\infty \chi(x;K_u)\,du,$$

where $K_u = \{x:\ f(x) \geqslant u\}$ is a symmetric convex set in R^n. Following this type of decomposition, Kanter defined a random vector in R^n to be symmetric unimodal, if its distribution is a "mixture" (with respect to a probability measure) of all uniform probability distributions on symmetric compact convex sets in R^n. It has been shown by Kanter that his symmetric unimodal functions are closed under weak convergence, and so they are essentially central convex UM.

Since log-concavity of measures (or densities) is closed under convolution, it follows easily that the class of symmetric unimodal functions of Kanter is closed under convolution [10]. It is still an open question whether monotone unimodality is closed under convolution.

Sherman [15] conjectured that a monotone UM distribution in R^n is in the closed (in L_1-norm) convex hull of all uniform distributions on symmetric compact convex sets in R^n. However, using an example of Dharmadhikari and Jogdeo [6], Wells [18] has shown that a monotone UM distribution in R^2 need not be central convex UM.

3. Questions on Marginal Functions

The basic question here is whether a marginal of a symmetric unimodal function is unimodal. Das Gupta [2] has shown that a marginal function of a symmetric unimodal function is ray-unimodal, but such a marginal function may fail to satisfy Anderson's condition for unimodal functions. To prove Das Gupta's first result it is sufficient to consider the indicator function of a symmetric compact convex set C in the space of x and y, $x \in R^n$, $y \in R^m$. Let

$$C(y) = \{x \in R^n: (x,y) \in C\}.$$

Note that

$$C(\lambda_1 y_1 + \lambda_2 y_2) \supset \lambda_1 C(y_1) + \lambda_2 C(y_2).$$

It now follows from Brunn-Minkowski inequality that

$$\mu_n[C(\lambda_1 y_1 + \lambda_2 y_2)] \geqslant \min[\mu_n(C(y_1)), \ \mu_n(c(y_2))),$$

where $0 \leqslant \lambda_1, \lambda_2 \leqslant 1$, and μ_n is the Lebesgue measure on R^n. Specializing $\lambda_1 = (1+\lambda)/2$, $y_1 = y$, $y_2 = -y$, and noting that $C(y) = -C(-y)$, we get

$$\mu_n(C(\lambda y)) \geqslant \mu_n(C(x)).$$

Anderson's inequality follows from Das Gupta's result by considering the function h defined by

$$h(x,y) = f(x+y) \ \chi(x;E).$$

Furthermore, Das Gupta [2] has shown that a marginal of the product of k symmetric unimodal functions is ray-unimodal. Dharmadhikari and Jogdeo [6] have shown that both central convex unimodality and monotone unimodality are inherited by marginal functions. A similar result also holds for symmetric unimodal functions of Kanter [10].

4. Results on Log-concave Functions

If the function f in Anderson's Theorem happens to be log-concave or strongly unimodal and E is a convex set in R^n, then Prekopa's Theorem implies that $\int_E f(x+y)dx$ is a log-concave function of y. Prekopa's Theorem [14] states that the convolution of two log-concave functions is log-concave. This again is a consequence of the fact that a marginal function of a log-concave function is log-concave. The above result on convolution was proved by Davidovic, Karenbljum and Hacet [5] using a weaker version of Anderson's inequality. The fact that marginality preserves the log-concavity property follows from Das Gupta's result. The key to the proof of this result is the following. If g is a log-concave function defined on $R^n \times R^m$, then

$$f(y,v;x,u) \equiv g(x-y,(u-v)/2)\ g(x+y,(u+v)/2)$$

is a centrally symmetric unimodal function in (y,v) for every (x,u); $x,y \in R^n$, $u,v \in R^m$. On the other hand, the above result also implies Brunn-Minkowski inequality which was used to prove Anderson's inequality. To see this fact note that for any two convex sets A_0 and A_1 in R^n the characteristic function of the set

$$D = \{(\theta,x):\theta \in [0,1],\ x \in (1-\theta)A_0 + \theta A_1\}$$

is strongly unimodal. Next note that

$$(1-\theta)A_0 + \theta A_1 = [(1-\eta)A_0^* + \eta A_1^*]\ [(1-\theta)\mu_n^{1/n}(A_0) + \theta\mu_n^{1/n}(A_1)],$$

where

$$\eta = \theta\mu_n^{1/n}(A_1)/[(1-\theta)\mu_n^{1/n}(A_0) + \theta\mu_n^{1/n}(A_1)],$$
$$A_i^* = A_i/\mu_n^{1/n}(A_i).$$

5. More General Invariance and Pre-ordering

If f is a unimodal function on R^n, invariant under a group G of Lebesgue-measure preserving transformations, and E is a G-invariant convex set in R^n, then

$$\int_E f(x-y^*)\,dx \geqslant \int_E f(x-y)\,dx$$

where y^* lies in the convex hull of the G-orbit of $\{y\}$. Anderson's inequality is the special case of the above result, when G is the group of sign transformations. The above generalization is due to Mudholkar [13]. To see this result, specialize $y_1 = g_1 y$ and $y_2 = g_2 y$ in (2.3), where g_1 and g_2 are elements in G. Next note that

$$h(y,u) = h(gy,u)$$

for all g in G. A similar generalization for marginal functions has been obtained by Das Gupta [2].

Mudholkar's generalization led to an interesting development as follows. Let C(y) be the convex-hull of the G-orbit of $\{y\}$. If we write $y^* \leqslant y$ if $y^* \in C(y)$, then it follows easily that $\leqslant$ is a pre-order on R^n. We call a function h on R^n G-decreasing if $y^* \leqslant y$ implies $h(y^*) \geqslant h(y)$. It is easy to see that if f is a G-invariant unimodal function then f is G-decreasing.

Mudholkar's result naturally led to the following question [7]: For which groups G is the function defined by

$$h(y) = \int f_1(x) f_2(x-y)\,dx$$

would be G-decreasing, if f_1 and f_2 are non-negative and G-decreasing? It is now known that the above result holds when G is the permutation group [12] or more generally the reflection groups [8].

Incidentally, these results also use Anderson's inequality as the basic step. One of the key facts is the following: For a non-negative G-decreasing function h on R^n with G as the permutation group

$$h(u+v,\ u-v,\ x_3,\ldots,x_n)$$

is centrally symmetric unimodal function of v only.

6. Ordering of Distributions

As a corollary to his Theorem, Anderson [1] proved the following result: If $X \sim Np(0,\Sigma)$, $Z \sim Np(0,\Gamma)$ and $\Gamma - \Sigma$ is positive semi-definite, then for any symmetric convex set in R^p

$$P(X \in C) \geqslant P(Z \geqslant C). \qquad (6.1)$$

This is an easy consequence of Anderson's inequality, since the normal density with zero means is symmetric and unimodal and Z can be expressed as $Z = X + Y$ where $Y \sim Np(0,\ \Gamma - \Sigma)$ independently of X.

However, the relation (6.1) may be used to define an ordering (with respect to more concentration about 0) between two distributions. More generally, one may write $P_1 < P_2$ for two distributions P_1 and P_2 iff

$$\int f(x)dP_1(x) \geqslant \int f(x)dP_2(x)$$

for all functions f in the convex cone generated by the indicators of congex symmetric sets [7].

7. Results on Association

Although Anderson's result deals with translation shift it has been used to derive interesting results on association and correlation. Let us go back to Anderson's theorem and assume that the function H in (2.1) is differentiable. Furthermore, assume that the differentiation can be done within the sign of integral. Since H(ky) is a nonincreasing function of $k > 0$, differentiating H(ky) with respect to k we set

$$(7.1) \qquad \int_E \sum_{i=1}^{n} y_i \frac{\partial f(x+ky)}{\partial x_i} \leqslant 0$$

for $k > 0$, where $y = (y_1 \ldots y_n)$, $x = (x_1 \ldots x_n)$.

The above relation (7.1) is used to derive the following result:

Theorem. Let $(x_1,\ldots,x_n)$ be jointly normally distributed with zero means and covariance matrix $\Sigma = (\sigma_{ij})$. Let $\Sigma(\lambda)$ be the covariance matrix with $\sigma_{1j}(\lambda) = \lambda\sigma_{1j}$ for $j > 1$, and $\sigma_{ij}(\lambda) = \sigma_{ij}$ for all other i and j; $0 \leqslant \lambda \leqslant 1$. Let P_λ be the normal distribution with zero means and covariance matrix $\Sigma(\lambda)$. Then for every $C_1 > 0$ and symmetric convex set C_2 in R^{n-1},

$$(7.2) \qquad P_\lambda[|X_1| \leqslant C_1,\ (X_2,\ldots,X_n) \in C_2]$$

is nondecreasing in $\lambda \in [0,1]$.

The above result is due to Sidak [16], and Sidak's proof was simplified by Jogdeo [9]. This result has been extended to elliptically contoured symmetric distributions by Das Gupta et al. [ ]. A particular case of the above Theorem is the following:

$$(7.3) \qquad P[|X_i| \leqslant C_i, i = 1,\ldots,n] \geqslant \prod_{i=1}^{n} P(|X_i| \leqslant C_i).$$

More generally, one may consider the following probability:

$$\pi(\lambda) = P_\lambda[X_{(1)} \in C_1,\ X_{(2)} \in C_2].$$

where $X = (X_{(1)}, X_{(2)})$, C_i is a symmetric convex set in the space of $X_{(i)}$, and P_λ refers to the normal distribution of X with zero means and covariance matrix $\Sigma(\lambda)$ given by

$$\Sigma(\lambda) = \begin{bmatrix} \Sigma_{11} & \lambda\Sigma_{12} \\ \lambda\Sigma_{21} & \Sigma_{22} \end{bmatrix} ,$$

$0 \leqslant \lambda \leqslant 1$. Pitt [17] has shown that $\pi(\lambda)$ is an increasing function of λ when rank $(\Sigma_{12}) \leqslant 2$. Khatri [11] has proved earlier that

$$P[X_{(1)} \in C_1,\ X_{(2)} \in C_2] \geqslant P[X_{(1)} \in C_1]\ P[X_{(2)} \in C_2]$$

when rank $(\Sigma_{12}) = 1$.

Pitt's proof uses the fact that the marginal of a log-concave function is log-concave. On the other hand, Khatri's proof depends more directly on Anderson's inequality. All the above results have been proved by using a conditional argument and the relation (7.1) (or, the original form of Anderson's inequality).

Remark. Anderson's inequality has been applied extensively to get many important results on power functions of multivariate tests, confidence regions, and association of random variables. However, in this review we have tried to restrict our attention only to probability inequalities.

References

1. Anderson, T.W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. Proc. Amer. Math. Soc. **6**, 170-176.

2. Das Gupta, S. (1976). A generalization of Anderson's Theorem on unimodal functions. Proc. Amer. Math. Soc. **60**, 85-91.

3. Das Gupta, S., Eaton, M.L., Olkin, I., Perlman, M.D., Savage, L.J., and Sobel, M. (1970). Inequalities on the probability content of convex regions for elliptically contoured distributions. Proc. VI Berkeley Symp. on Probability and Statistics, Vol. II, Univ. of Calif., Berkeley.

4. Das Gupta, S. (1980). Brunn-Minkowski Inequality and its Aftermath. J. Multiv. Anal., **10**, 296-318.

5. Davidovic, Ju.S., Korenbljum, B.I. and Hacet, R.I. (1962). A property of logarithmically concave functions. Soviet Math. Dok. **10**, No. 2, 477-480.

6. Dharmadhikari, S.W. and Jogdeo. K. (1976). Multivariate Unimodality. Ann. Statist. **4**, 607-613.

7. Eaton, M.L. (1982). A Review of Selected Topics in Multivariate Probability Inequalities, Ann. Statist. **10**, 11-43.

8. Eaton, M.L. and Perlman, M.D. (1977). Reflection groups, generalized Schur functions and the geometry of majorizations. Ann. Statist., **5**, 829-860.

9. Jogdeo, K. (1977). Association and Probability Inequalities. Ann. Stat. **5**, 495-504.

10. Kanter, M. (1977). Unimodality and dominance for symmetric random vectors. Trans. Amer. Math. Soc. **229**, 65-85.

11. Khatri, C.G. (1967). On certain inequalities for normal distributions and their applications to simultaneous confidence bounds. Ann. Math Statist. **38**, 1853-1867.

12. Marshall, A.W. and Olkin, I. (1974). Majorization in multivariate distributions. Ann. Statist. **2**, 1189-1200.

13. Mudholkar, G.S. (1966). The integral of an invariant unimodal function over an invariant convex set-an inequality and applications. Proc. Amer. Math. Soc. **17**, 1327-1333.

14. Prekopa, A. (1973). On logarithmic concave functions. Acta. Sci. Mat. **34**, 335-343.

15. Sherman, S. (1955). A Theorem on convex sets with applications. Ann. Math. Statist. **26**, 763-766.

16. Sidak, Z. (1968). On multivariate normal probabilities of rectangles; their dependence on correlations. Ann. Math. Statist. **39**, 1425-1434.

17. Pitt, L. (1977). A Gaussian correlation inequality for symmetric convex sets. Ann. Probability **5**, 470-474.

18. Wells, D.R. (1978). A result in multivariate unimodality. Ann. Statist. **6**, 926-931.